March 1966

BEST SEQUENTIAL TESTS FOR AN IDENTIFICATION PROBLEM WITH PARTIALLY OVERLAPPING DISTRIBUTIONS

by

Saul Blumenthal[1], Theodore Christie, and Milton Sobel[2]

Rutgers University, Rutgers University,

and University of Minnesota

Technical Report No. 74

University of Minnesota

Minneapolis, Minnesota

[1] Research supported in part by National Science Foundation Grant No. GP4933.

[2] Research supported by National Science Foundation Grant No. GP3813

1. Introduction

We are given k (≥2) populations from which observations can be taken, and two completely specified distribution functions $F_1(x)$ and $F_2(x)$. It is assumed that the observations from (k-1) of the populations have distribution $F_1(x)$ and that those from the remaining population have distribution $F_2(x)$. The problem considered is that of identifying the population associated with $F_2(x)$, subject to certain restrictions on F_1 and F_2 which are outlined below.

Let $f_i(x)$ (i=1,2) denote the corresponding density functions if both $F_1(x)$ and $F_2(x)$ are continuous or the probability masses at x if both $F_1(x)$ and $F_2(x)$ are discontinuous. Let

$$\Omega_i = \{x : f_i(x) > 0\} \qquad i = 1,2 \tag{1.1}$$

and let

$$\Omega = \Omega_1 \cup \Omega_2 \tag{1.1a}$$

We shall treat the "degenerate" situation in which the likelihood ratio $\left(f_2(x)/f_1(x)\right)$ assumes only the three values $(0,1,\infty)$ for $(x \in \Omega)$. That is, on $(\Omega_1 \cap \Omega_2)$, $f_1(x) = f_2(x)$, and

$$P_{F_1}(\Omega_1 \cap \Omega_2) = P_{F_2}(\Omega_1 \cap \Omega_2) = \theta. \quad (0 < \theta < 1). \tag{1.2}$$

A typical example of such a configuration has $F_1(x)$ assigning mass $(1-\theta)$ to $x = -1$ and mass θ to $x = 0$, while $F_2(x)$ assigns mass θ to $x = 0$ and mass $(1-\theta)$ to $x = 1$ (two symmetrically degenerate trinomial distributions). Another important example occurs with $F_1(x)$ as the uniform distribution on $(\alpha, \alpha + 1)$ and $F_2(x)$ as the uniform distribution on $(\alpha + \theta, \alpha + \theta + 1)$. Note that since α must be specified, this cannot be formulated as a slippage problem or a ranking problem for the uniform distribution.

An identification is an association of one of the populations with the distribution $F_2(x)$. We shall consider sequential procedures which satisfy the requirement

$$(1.3) \qquad \min P\ (\text{Correct Identification}) \geq P^* ,$$

where the minimum is taken over the k possible true associations of distribution $F_2(x)$ with a population, and P^* $(1/k < P^* \leq 1)$ is a specified constant. Among these procedures we seek those rules T^* which satisfy

$$(1.4) \qquad \max E(N|T^*) = \min_{T} \max E(N|T) ,$$

where the maximum is taken over the k possible true pairings, the rules T all satisfy (1.3), and N is the total number of observations taken.

Because of the degeneracy previously described, when $P^* = 1$ it is possible to satisfy requirement (1.3) using a sampling procedure whose sample size distribution has finite moments of all orders. When $P^* < 1$, it is necessary to introduce randomization to achieve equality in (1.3). We shall show that an entire family of randomized procedures achieves the requirements (1.3) and (1.4). From this family we find one procedure having minimum variance of the sample size. This procedure is "truncated" rather than fully randomized, the method of truncation being described in Section 2.

In Section 2, some general discussion is given in which we limit the class of procedures which need be considered, and the proposed procedures are described. The standard identification procedure of [1] is also described for comparison. In Sections 3 and 4, best vector-of-observations and one observation at a time procedures are derived for the case of two populations. In Section 5 these are compared with one another and with some sub-optimal procedures. In Section 6, special properties of procedures achieving $P^* = 1$ for general k are considered. In Sections 7 and 8 the best vector and one observation at a time procedures are derived for general k, and in Section 9 these are compared with one another and with a suitable single-sample procedure.

2. Generalities and Description of Procedures

Denote the k populations by $(\pi_1,\pi_2,\ldots,\pi_k)$, and the j^{th} random variable from π_i by X_{ij}, $i = 1,2,\ldots,k$; $j = 1,2,\ldots$. The X_{ij} are assumed to be independent. Define the likelihood of the sample $\{x_{ij}\}$ $(i = 1,\ldots,k;\ j = 1,\ldots,m)$ given the true association (π_i, F_2) as

$$(2.1)\qquad L(i,m) = \prod_{j=1}^{m} \left\{ f_2(x_{ij}) \prod_{\substack{s=1 \\ s\neq i}}^{k} f_1(x_{sj}) \right\},\ i = 1,\ldots,k$$

and the i^{th} likelihood ratio as

$$(2.2)\qquad R(i,m) = L(i,m) \Big/ \sum_{s=1}^{k} L(s,m)\ ,\ i = 1,\ldots,k.$$

Also, let

$$(2.3)\qquad R_m = \max_{1 \le i \le k} R(i,m)$$

If a-priori probability (1/k) is assigned to each of the k possible (π_i,F_2) pairings, then R(i,m) can be regarded as the a-posteriori probability of the pairing (π_i,F_2). Note that each R(i,m) (i = 1,...,k) assumes one of the values 0,1/k, 1/(k-1),...,1/2,1.

The procedure considered in [1], known as the basic identification procedure, uses the stopping rule: "continue sampling until $R_m \ge P^*$; then stop and make an identification". It is clear that such a procedure will satisfy (1.3), but in view of the previous paragraph, it is also clear that if $P^* > \frac{1}{2}$, the actual P(CI)(probability of correct identification) will be unity.

The following discussion is not specifically directed to the basic procedure.

Because of the degeneracy in this problem, the set of R(i,m) values for the sample $\{x_{im}\}$ will fall into one of the following two categories:

(i) $(k-1)$ of the $R(i,m)$ values are zero and the remaining one is unity

or (ii) $(k-C)$ $(2 \le C \le k)$ of the $R(i,m)$ values are zero and the remaining $R(i,m)$ values are $1/C$.

In case (i), it is known exactly which pairing is correct, while in case (ii) it is known that $(k-C)$ pairings are impossible but the sample contains no information to differentiate among the C possible pairings. Thus to achieve equality in (1.3) when $1/2 < P^* < 1$, a randomized stopping rule of some form must be used. (From the above considerations it is clear that to any rule which uses the data to decide whether to stop and to choose among the C "contenders", there corresponds a purely randomized rule which achieves the same P(CI) and the same distribution of sample size). Achievement of equality in (1.3) is desirable when $P^* < 1$ in order to reduce the average sample size from that required to achieve P(CI) = 1. It may at first seem unreasonable to accept a P(CI) of less than unity when a P(CI) of unity can be achieved using the basic identification procedure, but since no fixed sample size procedure can achieve a P(CI) of unity we do not require this of a sequential procedure at the expense of a large expected sample size.

It is evident that nothing can be learned by continuing to take observations from population Π_j if $R(j,m) = 0$, i.e., if some χ_{jk} $(k = 1,\ldots,m)$ falls in $(\Omega-\Omega_2)$ (see (1.1) and (1.2)). Thus we consider only sampling rules which stop sampling from a population Π_j as soon as $R(j,m)$ becomes zero (i.e., rules which take no observations from Π_j beyond stage m if $\chi_{j1},\ldots,\chi_{j(m-1)}$ are all in $(\Omega_1 \cap \Omega_2)$ and χ_{jm} is in $(\Omega-\Omega_2)$). A population for which $R(\cdot,m) = 0$ is called a "non-contender". The C populations for which $R(\cdot,m) = 1/C$ are called "contenders", and we shall write C as $C(m)$ to indicate that C is a (non-increasing) function of the number of elapsed sampling stages. Observations at stage $(m+1)$ are taken only from the $C(m)$ populations which remain in contention after stage m.

Another obvious reduction in the class of rules to be considered is obtained from the symmetry of the requirements (1.3) and (1.4). We need consider only rules which use the observations symmetrically and which, when and if they stop with more

than one population still in contention, assign equal probabilities to the identification of F_2 with each of the remaining contenders. Given any non-symmetric rule satisfying (1.3), the symmetrized version of that rule will also satisfy (1.3) and will have at least as small a max E(N), with strict inequality if the original rule did not have the same E(N) for all possible true pairings. Thus, with the given requirements the symmetric rules form an essentially complete class.

Two basic sampling plans will be studied:

(a) "vector at a time" procedures which at each sampling stage take one observation from each of the C contenders, and

(b) "one at a time" procedures, which at each sampling stage take an observation from one of the C contenders.

By a randomized rule, we mean one which if $C(m) > 1$ assigns a probability $p_m (m=0,1,2,\ldots)$ to carrying out an $(m+1)^{st}$ stage of sampling and probability $(1-p_m)$ to terminating the experiment without the $(m+1)^{st}$ sample, and whenever sampling is terminated by randomization assigns probability 1/C to each pairing of a contender with distribution F_2. The probability p_m may be allowed to depend on the entire sequence $\{x_{ij}\}$, $i=1,\ldots,k, j=1,\ldots,m$ of previous observations. (Randomized procedures which eliminate contenders without terminating sampling will not be discussed in this paper.) Thus $(1-p_o)$ is the probability of taking no observations at all, and if sampling is started it will be terminated at the first stage m for which C(m) is unity or else by the randomizing scheme described above.

It will be shown that when $k = 2$, all rules which achieve (1.3) with equality have the same value of E(N) (the maximum being omitted since the rules are symmetric). Among these rules, the one which minimizes $E(N^2)$ (and thus minimizes Var (N)) is a "truncated" rule which randomizes at at most its final stage to achieve equality in (1.3). Thus the "best" randomized rule in this sense is not really a fully randomized one.

When $k > 2$, rules which achieve (1.4) subject to (1.3) have the following form: if $C(m) = k$, a positive probability $(1-p_m)$ can be assigned to not taking an $(m+1)^{st}$

stage of observations, but if $2 \le C(m) \le k-1$, the probability of observing at stage $(m+1)$ is unity. Thus as soon as one non-contender is found, sampling is allowed to continue until it is known with certainty which population is paired with F_2. Any randomization takes place only at stages where all k populations are still in contention.

Whereas the rule for the case when $k = 2$ can be thought of as a natural intuitive generalization of the basic identification procedure of [1], the rule for $k > 2$ cannot. To illustrate the most striking difference, consider the case of $k = 4$ and $P^* = 0.40$. The basic procedure would stop as soon as two contenders were eliminated and would achieve a P(CI) of $(\frac{1}{2})$. The natural extension of this would be to randomize when there are three or four contenders, and stop when there are two. Proper choice of the randomizing probabilities will yield $P(CI) = 0.40$. The "best" rule described above, however, applies randomized stopping only when there are four contenders, and if one or more populations are eliminated from contention, continues sampling until only one contender remains. Thus the conditional P(CI) at stopping is either $(\frac{1}{4})$ or 1, never being $(\frac{1}{3})$ or $(\frac{1}{2})$, and an unconditional P(CI) of 0.40 is achieved by proper choice of the randomizing probabilities.

All rules of the above form which achieve (1.3) with equality have the same E(N), and the one with the smallest $E(N^2)$ is one which has all the stopping probabilities $(1-p_m)$ equal to zero for $m < n_o-1$ and equal to unity for $m \ge n_o$, with n_o and p_{n_o-1} chosen to achieve equality in (1.3). This is a "pseudo-truncated" rule in that if no contenders are eliminated by the n_oth stage, sampling is terminated, but if any contender is eliminated, no upper bound is set on the additional sampling needed to terminate. Again, the "best" randomized rule is not really randomized.

The only difference between the vector at a time and one at a time procedures is that at sampling stage $(m+1)$, the vector rule takes a sample of $C(m)$ observations, whereas the one at a time rule selects at random one of the $C(m)$ contenders from which to take an observation.

3. Two populations, vector sampling.

Let p_{m-1} $(0 \le p_{m-1} \le 1)$ denote the probability of taking an m^{th} stage of observation $(m=1,2,\ldots)$, and let α_{m-1} be the conditional probability of choosing pairing (Π_1, F_2) if sampling is terminated by randomization after stage $(m-1)$. The remarks of Section 2 indicate that we could restrict attention to $\alpha_m = \frac{1}{2}$, but we shall allow the full range, $0 \le \alpha_m \le 1$. Let N be the total number of observations up to stopping and let M be the number of sampling stages up to stopping (here $M = N/2$). Then for $m \ge 1$, we have using (1.2)

$$(3.1a) \qquad P(M \ge m) = \theta^{2(m-1)} p_0 p_1 \cdots p_{m-1}$$

$$(3.1b) \qquad P(M = m | M \ge m) = 1 - \theta^2 p_m$$

$$(3.1c) \qquad P\left(CI, M = m | M \ge m;\ (\Pi_1, F_2) \text{ is true}\right) = 1 - \theta^2 + \alpha_m(1 - p_m)\theta^2$$

$$(3.1d) \qquad P\left(CI, M = m | M \ge m;\ (\Pi_2, F_2) \text{ is true}\right) = 1 - \theta^2 + (1-\alpha_m)(1 - p_m)\theta^2$$

Theorem 3.1 Let $\frac{1}{2} < P^* \le 1$ and $\theta > 0$ be given. For any sequences $\{p_i\}$ $(i=0,1,2,\ldots)$ and $\{\alpha_i\}$ $(i=0,1,2,\ldots)$ for which (1.3) is satisfied we have that

$$(3.2) \qquad E(M) \ge (2P^* - 1)/(1-\theta^2)$$

for both true pairings, independent of the $\{p_i\}$ and $\{\alpha_i\}$, with equality if and only if

$$(3.3) \qquad P(CI) = P^*$$

for both true pairings.

Proof: From equations (3.1) it is seen that

$$(3.4a) \quad P\left(CI | (\Pi_1, F_2)\right) = (1-p_0)\alpha_0 + \sum_{m=1}^{\infty} P(M \ge m)\left[1 - \theta^2\{1 - \alpha_m(1-p_m)\}\right]$$

$$(3.4b)\quad P\left(CI|(\Pi_2,F_2)\right) = (1-p_0)(1-\alpha_0)+\sum_{m=1}^{\infty} P(M\geq m)\left[1-\theta^2\{1-(1-\alpha_m)(1-p_m)\}\right].$$

Equivalently, we write (3.4) as

$$(3.5a)\quad \left[P\left(CI|(\Pi_1,F_2)\right) + P\left(CI|(\Pi_2,F_2)\right)\right] = (1-p_0)+\sum_{m=1}^{\infty} P(M\geq m)\left[2-\theta^2(1+p_m)\right]$$

$$(3.5b)\quad \left[P\left(CI|(\Pi_1,F_2)\right) - P\left(CI|(\Pi_2 F_2)\right)\right] = (1-p_0)(2\alpha_0-1)+\sum_{m=1}^{\infty} P(M\geq m)\left[\theta^2(1-p_m)(2\alpha_m-1)\right].$$

Use of (3.1b), (3.1a), and simple manipulation gives for (3.5a)

$$(3.6)\quad \left[P\left(CI|(\Pi_1,F_2)\right)+P\left(CI|(\Pi_2,F_2)\right)\right] = 1+(1-\theta^2)\sum_{m=0}^{\infty} P(M\geq m+1)$$

But since M is a non-negative random variable, it is well known that

$$(3.7)\qquad E(M) = \sum_{m=0}^{\infty} P(M \geq m+1).$$

Hence,

$$(3.8)\quad E(M) =\left\{\left(\left[P\left(CI|(\Pi_1,F_2)\right) + P\left(CI|(\Pi_2 F_2)\right)\right]-1\right)/(1-\theta^2)\right\}\geq (2P^*-1)/(1-\theta^2)$$

the inequality following from (1.3). The statement (3.3) then follows directly from (3.8). This completes the proof.

Note that under (3.3), (3.5b) becomes

$$(3.9)\quad 0 = (1-p_0)(2\alpha_0-1) + \sum_{m=1}^{\infty} P(m\geq m)\left[\theta^2(1-p_m)(2\alpha_m-1)\right]$$

and (3.2) becomes

$$(3.10)\qquad E(M) = (2P^*-1) / (1-\theta^2).$$

The sequence $\{\alpha_i\}$ is restricted by (3.9), but subject to that restriction plays no role in determining E(M), as seen by examining (3.5a).

Among all procedures satisfying (3.9) and (3.10) we wish to find one which minimizes $E(M^2)$. To this end, we prove the next lemma.

Lemma 3.1 The earliest stage at which a sampling plan can be truncated subject to (3.3) is at stage r_o, given by

$$(3.11) \qquad r_o = \left\lceil \left(\log 2(1-P^*)\right) / 2 \log \theta) \right\rceil ,$$

where $\lceil \chi \rceil$ denotes the smallest integer greater than or equal to χ.

Proof: Truncation at stage r is equivalent to setting $p_j=0$, $j \geq r$, and subject to (3.3), (3.8) becomes

$$(3.12) \qquad 2P^*-1 \leq (1-\theta^2) \sum_{m=0}^{r-1} P(M \geq m+1) = (1-\theta^2) \sum_{m=0}^{r-1} \theta^{2m} \left(\prod_{j=o}^{m} p_j \right).$$

The smallest value of r satisfying (3.12) will be found by setting $p_o=p_1=\ldots=p_{r_o-1}=1$, giving r_o as $\lceil r \rceil$, where r is the solution of

$$(3.13) \qquad 2P^*-1 = (1-\theta^2) \sum_{m=o}^{r-1} \theta^{2m} = 1-\theta^{2r}.$$

This yields (3.11).

Using the lemma, we can now find the min $E(M^2)$ procedure.

Theorem 3.2 Let r_o be given by (3.11). Among all procedures satisfying (3.3), $E(M^2)$ is minimized by:

$$p_o = p_1 = \ldots = p_{r_o-2} = 1$$

$$(3.14) \qquad p_{r_o-1} = \left\{2(P^*-1) + \theta^{2(r_o-1)}\right\} / \left\{(1-\theta^2)\theta^{2(r_o-1)}\right\}$$

$$p_{r_o} = p_{r_o+1} = \ldots = 0 .$$

Proof: From (3.1a) we have

Among all procedures satisfying (3.9) and (3.10) we wish to find one which minimizes $E(N^P)$. To this end, we prove the next Lemma.

Lemma 3.1 The earliest stage at which a sampling plan can be truncated subject to (3.5) is at stage r_0, given by

$$(3.11)\qquad r_0 = \left\lceil \left(\log 2(1-\alpha)\right) / 2 \; [illegible] \right\rceil .$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Proof: Truncation at stage r is equivalent to setting $p_{r-1}^0 = 1$, and subject to (3.7), (3.8) becomes

$$(3.12)\qquad [illegible]$$

The smallest value of r satisfying (3.12) will be found by setting [illegible] giving r_0 as $\lceil r \rceil$, where r is the solution of

$$(3.13)\qquad [illegible]$$

This yields (3.11).

Using the lemma, we can now find the min $E(N^P)$ procedure.

Theorem 3.2 Let r_0 be given by (3.11). Among all procedures satisfying (3.7), $E(N^P)$ is minimized by:

$$p_0^0 = p_1^0 = \cdots = p_{r_0-2}^0 = 1$$

$$(3.14)\qquad p_{r_0-1}^0 = [illegible]$$

$$p_{r_0}^0 = p_{r_0+1}^0 = \cdots = 0 .$$

Proof: From (3.14) we have

$$(3.15) \qquad E(M^2) = \sum_{m=1}^{\infty} m^2 P(M \geq m)[1-\theta^2 p_m]$$

$$= \sum_{m=0}^{\infty} (m+1)^2 P(M \geq m+1) - \sum_{m=0}^{\infty} m^2 P(M \geq m+1)$$

$$= 2\sum_{m=0}^{\infty} mP(M \geq m+1) + \sum_{m=0}^{\infty} P(M \geq m+1) .$$

Using (3.6), the quantity to be minimized is

$$(3.16) \qquad \Sigma mP(M \geq m+1) ,$$

subject to the following restriction obtained from (3.1a):

$$(3.17) \qquad 0 \leq P(M \geq m+1) \leq \theta^{2m}.$$

Writing

$$(3.18) \qquad C_m = \{(1-\theta^2)P(M \geq m+1)\} / (2P^*-1) \qquad m = 0,1,2,\ldots$$

the problem is to minimize $\sum_{m=0}^{\infty} mC_m$, subject to

$$(3.19) \qquad \begin{aligned} &0 \leq C_m \leq \theta^{2m}(1-\theta^2)/(2P^*-1) \\ &\sum_{m=0}^{\infty} C_m = 1. \end{aligned}$$

In this form, the C_m are probabilities to be chosen to minimize the expectation of the random variable Y, where $P(Y = m) = C_m$. Equation (3.14) follows immediately, completing the proof.

Note that for this best rule, only r_o, α_{r_o-1} and α_{r_o} need be determined, and this can be done using (3.9), which gives

$$(3.20) \quad 2\left\{\alpha_{r_o-1}(1-p_{r_o-1}) + \theta^2 p_{r_o-1}\alpha_{r_o}\right\} = 1-(1-\theta^2)p_{r_o-1} .$$

This does not determine the α's uniquely. The choice of $(\frac{1}{2})$ for both values does satisfy (3.20), and this corresponds to the rule described in Section 2.

4. Two populations, one at a time sampling

The notation used here is that of Section 3, with the additional representation that η_m $(0 \le \eta_m \le 1)$ is the conditional probability that the m^{th} observation is to be taken from population Π_1, given that sampling is still going on by the m^{th} stage. A procedure with an equivalent sample size distribution is one which before sampling begins assigns probability η to sampling from Π_1 and probability $(1-\eta)$ to sampling from Π_2, and then takes all observations from the "winning" population. For one at a time sampling $N \equiv M$. We take $\alpha_m = (\frac{1}{2})$ for notational convenience.

Analagous to equations (3.1) we have

$$(4.1a) \qquad P(N \ge m) = \theta^{m-1} p_o p_1 \cdots p_m$$

$$(4.1b) \qquad P(N = m | N \ge m) = 1-\theta p_m$$

$$(4.1c) \qquad P\left(CI, N=m | N \ge m, (\Pi_i, F_2)\right) = 1-\theta+(\tfrac{1}{2})(1-p_m)\theta \ , \quad i = 1,2.$$

Theorem 4.1 Let $\frac{1}{2} < P^* \le 1$ and $\theta > 0$ be given. For any sequences $\{p_i\}$ and $\{\eta_i\}$ (and $\{\alpha_i\}$) satisfying (1.3), we have

$$(4.2) \qquad E(N) \ge (2P^*-1) / (1-\theta)$$

for both pairings, independent of the sequences $\{p_i\}$ and $\{\eta_i\}$ (and $\{\alpha_i\}$), with equality if and only if (3.3) holds.

Proof: From (4.1c), (4.1a), and (3.7) we have

$$(4.3) \qquad \begin{aligned} P(CI) &= (\tfrac{1}{2})(1-p_o) + \sum_{m=1}^{\infty} P(N \ge m)\left(1-\theta+(\theta/2)(1-p_m)\right) \\ &= (\tfrac{1}{2}) + \left[(1-\theta)/2\right] \sum_{m=1}^{\infty} P(N \ge m) = (\tfrac{1}{2}) + \left[(1-\theta)E(N)\right] / 2. \end{aligned}$$

The Theorem follows as in the proof of Theorem 3.1.

In a result analogous to Theorem 3.2, we now give the procedure which minimizes $E(N^2)$ subject to (3.3).

Theorem 4.2 Among all one at a time procedures satisfying (3.3), $E(N^2)$ is minimized by the procedure with:

$$p_o = p_1 = \dots = p_{r_1-2} = 1$$

$$p_{r_1-1} = \left\{2(P^*-1) + \theta^{r_1-1}\right\} / (1-\theta)\theta^{r_1-1} \tag{4.4}$$

$$P_{r_1} = P_{r_1+1} = \dots = 0 ,$$

where r_1 is given by

$$r_1 = \left[\left(\log 2(1-P^*)\right) / \log \theta \right] \tag{4.5}$$

and $[x]$ denotes the smallest integer greater than or equal to x.

The proof is omitted, but note that the sequence $\{\eta_m\}$ or the value η is totally irrelevant when there are only two populations.

5. Comparison of rules, two populations

Although the average sample size for the one at a time procedure $\left(E_o(N)\right)$ is smaller than that of the vector at a time procedure $\left(E_v(N)\right)$, the difference is in all cases relatively small. From Theorems 3.1 and 4.1, it can be seen that

$$\left(E_v(N) / E_o(N)\right) = 2/(1+\theta) \tag{5.1}$$

which is 2 at $\theta = 0$ and 1 at $\theta = 1$ (identical populations).

Also,

$$E_v(N) - E_o(N) = (2P^*-1) / (1 + \theta) \tag{5.2}$$

which ranges from $(2P^*-1)$ at $\theta = 0$ to $(2P^*-1)/2$ at $\theta = 1$ (when both expectations become infinite). Note that this difference never exceeds unity. Thus the average saving in sample size by one at a time sampling is small, and the average number of stages of sampling is almost twice that of the vector at a time procedure. If sampling cost is more a function of the number of stages than of observations, the vector procedure could be more economical.

From Lemma 3.1 and Theorem 4.2, it can be seen that both procedures have the same maximum sample size, with the maximum number of stages for the vector procedure again being about half that of the other procedure.

If we consider the single sample procedure which takes R pairs of observations (X_{1i}, X_{2i}) $(i=1,\dots,R)$, chooses pairing $(\pi_{1(2)}F_2)$ if for some i

$$X_{2(1)i}\,\varepsilon(\Omega_1 - \Omega_2) \text{ or } X_{1(2)}\,\varepsilon(\Omega_2 - \Omega_1)\ ,$$

and assigns probability $(\frac{1}{2})$ to each pairing if none of these events occurs, it is easily seen that R is given by (3.11). Thus the sequential procedure never takes more observations than the single sample procedure, and could be viewed as a "curtailed" single sample rule.

To see how the variance of N can be affected by the choice of the $\{p_j\}$ among sequential procedures satisfying (3.3), consider the very simple randomized rule which has $p_j = p$, $j=1,2,\dots$. It is easily seen that

$$(5.3) \qquad p = (2P^*-1) \Big/ \left[1-2\theta^2(1-P^*)\right]$$

and

$$(5.4) \qquad E(N^2) = (2P^*-1)\left[3(1-\theta^2) + 2\theta^2(2P^*-1)\right] \Big/ (1-\theta^2)^2\ .$$

For the best procedure, using (3.14) and (3.15) and assuming that $\left(\log 2(1-P^*)/2 \log \theta\right.$ is an integer, we have

$$(5.5) \qquad E(N^2) = (2P^*-1)(1+\theta^2)/(1-\theta^2)^2\ .$$

The difference between (5.4) and (5.5) is

$$(5.6) \qquad (2P^*-1)\left[2-4\theta^2(1-P^*)\right] / (1-\theta^2)^2 ,$$

which ranges from $2(2P^*-1)$ at $\theta = 0$ to infinity at $\theta = 1$.

6. General k, $P^* = 1$

Before proceeding with a discussion of randomized rules to achieve a P^* value of less than unity, we consider the case where randomization cannot be used, i.e., $P^*=1$. Letting $C(m)$ be the number of contenders after sampling at stage m, note that the sequence $\{C(m)\}$ $m=0,1,\ldots$ forms a sequence of state variables for a stationary finite Markov chain. The states of the chain are $(1,\ldots,k)$, representing the number of contenders; state 1 is "absorbing", corresponding to the identification of the correct population associated with F_2, while the states $2,\ldots,k$ are transient. Stationarity is due to the fact that no randomization is used so that transition probabilities depend only on the value of $C(m)$. Transitions can occur only to a lower numbered state. For vector sampling, all states with indexes no greater than $C(m)$ are possible at stage $(m+1)$, while for one at a time sampling the only possible transitions are to $C(m)$, $\big(C(m)-1\big)$ or 1. The number of sampling stages of the plan, M, corresponds to the number of steps until absorbtion for the Markov chain, while the number of observations, N, of the plan corresponds to the value of the chain functional

$$(6.1) \qquad G = \sum_{m=1}^{M} C(m) .$$

Having made the correspondence between the pertinent quantities for the sampling scheme and the appropriate Markov chain quantities, we shall not use results from Markov theory to find these quantities, but shall use direct arguments which then yield formuli for the Markov chain.

For the vector procedure, the "transition" probabilities are given by

$$(6.2a) \quad P(i,j) = P\left\{C(m+1) = j \mid C(m) = i\right\} = \binom{i-1}{i-j}\theta^j(1-\theta)^{i-j} \quad 2 \le j \le i \le k$$

(6.2b) $P(i,1) = P\{C(m+1) = 1 \mid C(m) = i\} = \theta(1-\theta)^{i-1} + (1-\theta) \quad 2 \le i \le k$.

Another random variable of interest is $S^{(i)}$, the number of steps spent in state i $(2 \le i \le k)$ given that state i has been entered.

We introduce the notation $E(X|k,P)$ to represent the expected value of X when there are k populations and the probability of correct identification is P - anticipating the next section. Here we take P = 1. We now state as a lemma a few obvious conclusions from the above definitions.

<u>Lemma 6.1</u> For the vector sampling procedure,

(6.3a) $E(S^{(i)}|k,1) = 1/(1-\theta^i) \qquad 2 \le i \le k$,

(6.3b) $E(N|k,1) = (1-\theta^k)^{-1} \left\{k + \sum_{j=2}^{k-1} P(k,j)E(N|j,1)\right\}$,

(6.3c) $E(M|k,1) = (1-\theta^k)^{-1} \left\{1 + \sum_{j=2}^{k-1} P(k,j)E(M|j,1)\right\}$,

(6.3d) $E(N|2,1) = 2E(M|2,1) = 2(1-\theta^2)^{-1}$,

(6.3e) $E(N|k,1) = A_k(I-P)^{-1}B_k$,

and

(6.3f) $E(S|k,1) = A_k(I-P)^{-1}D_k$,

where $A_k' = (1,0,0\ldots,0)$, $B_k' = (k,k-1,\ldots,2,0)$ and $D_k' = (1,\ldots,1)$ are k-element row vectors, I is the kxk identity matrix, and P is the kxk transition matrix $\{P(i,j)\}$ with elements given by (6.2a) and (6.2b).

<u>Proof:</u>

Equation (6.3a) follows from the fact that $S^{(i)}$ has a geometric distribution with parameter θ^i, while (6.3b) and (6.3c) are straightforward recursions, (6.3d) restates (3.10), while (6.3e) and (6.3f) follow from the usual arguments to be found, for instance, in Kemeny and Snell [2]. The vector A_k gives the "initial" state probabilities.

We now obtain explicit formulas for $E(N|k,1)$ and $E(M|k,1)$ which can be regarded as computing formulas for the matrix expressions in (6.3e) and (6.3f), which are not easily evaluated by direct means.

Theorem 6.1 The expected number of observations for vector at a time sampling when $P^*=1$ is given by

$$(6.4)\qquad E(N|k,1) = [2(k-1)/(1-\theta^2)] + \sum_{j=3}^{k} \left[(-1)^j \binom{k-1}{j-1} / (1-\theta^j)\right]$$

Proof: Assume for concreteness that pairing (Π_k, F_2) is correct. Let $T_i (i=1,\ldots,k)$ be the number of observations taken from population Π_i. Let Y_i represent the number of observations from Π_i until the first time an observation falls in $(\Omega_1 - \Omega_2)$ $(i=1,\ldots,k-1)$ or in $(\Omega_2 - \Omega_1)(i=k)$. Clearly $Y_i \geq T_i$, all i, and the Y_i are independent geometrically distributed variables with parameter θ. Let

$$(6.5)\qquad \hat{Y} = \max\,(Y_1,\ldots,Y_{k-1})$$

be the index at which all the (k-1) incorrect pairings have been eliminated. Since sampling stops when all incorrect pairings are eliminated or when the correct pairing is discovered (at index Y_k) the stage number at which sampling ceases is

$$(6.6)\qquad Z = \min\,(Y_k, \hat{Y}) .$$

Also, since sampling from population Π_i ceases as soon as Π_i is eliminated (at Y_i) or the true pairing discovered (at Y_k), the number of observations taken from Π_i is given by

$$(6.7)\qquad T_i = Y_i(Y_k) \qquad i = 1,\ldots,k-1$$

where by $Y_i(z)$ we mean the random variable Y_i curtailed at the integer z, i.e.

(6.8)
$$P\{Y_i(z) = j\} = P\{Y_i = j\} = \theta^{j-1}(1-\theta) \text{ for } j = 0,1,\ldots,z-1$$
$$P\{Y_i(z) = z\} = P\{Y_i \geq z\} = \theta^{z-1} .$$

The desired quantity, N, is then

$$N = \sum_{i=1}^{k-1} T_i + Z \tag{6.9}$$

and

$$E(N|k,1) = \sum_{i=1}^{k-1} E(T_i) + E(Z) . \tag{6.10}$$

For fixed $Y_k = y$, we find easily

$$E(T_i|Y_k=y) = E\Big(Y_i(y)\Big) = (1-\theta)\sum_{j=1}^{y-1} j\theta^{j-1} + y\theta^{y-1} = (1-\theta^y)/(1-\theta) \qquad (i = 1,\ldots,k-1) \tag{6.12}$$

so that noting the geometric distribution of Y_k, we have

$$E(N|k,1) = E(Z) + (k-1)\sum_{y=1}^{\infty}(1-\theta)\theta^{y-1}[(1-\theta^y)/(1-\theta)] \tag{6.13}$$

$$= E(Z) + [(k-1)/(1-\theta^2)] .$$

It is easily verified that

$$P(\hat{Y} \leq y) = (1-\theta^y)^{k-1} \tag{6.14}$$

and thus

$$P(Z \leq z) = P(\hat{Y} \leq z) + P(X_k \leq z) - P(\hat{Y} \leq z)P(X_k \leq z) \tag{6.15}$$

$$= (1-\theta^z)^{k-1} + (1-\theta^z) - (1-\theta^z)^k ,$$

so that

$$(6.16)\qquad P(Z=z) = \theta^{z-1}(1-\theta) + (1-\theta^z)^{k-1} - (1-\theta^{z-1})^{k-1} + (1-\theta^{z-1})^k - (1-\theta^z)^k \,,$$

$$z = 1,2,\ldots$$

We then obtain

$$(6.17)\quad E(Z) = (1-\theta)^{-1} + \sum_{z=1}^{\infty} z[(1-\theta^{z-1})^k - (1-\theta^z)^k] - \sum_{z=1}^{\infty} z[(1-\theta^{z-1})^{k-1} - (1-\theta^z)^{k-1}]$$

$$= (1-\theta)^{-1} + \sum_{z=1}^{\infty} [(1-\theta^{z-1})^k - (1-\theta^{z-1})^{k-1}]$$

$$= (1-\theta)^{-1} - \sum_{z=0}^{\infty} \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \theta^{z(j+1)}$$

$$= (1-\theta)^{-1} + \sum_{i=1}^{k}\left[\binom{k-1}{i-1}(-1)^i/(1-\theta^i)\right] .$$

Combining (6.17) and (6.13) gives (6.4). This completes the proof.

<u>Corollary 6.1</u> The expected number of stages until stopping for vector sampling is

$$(6.18)\quad E(M|k,1) = (1-\theta)^{-1} + \sum_{i=1}^{k}\left[\binom{k-1}{i-1}(-1)^i/(1-\theta^i)\right] .$$

<u>Proof</u>: Note that $M = Z$ as given in (6.6).

<u>Corollary 6.2</u> An upper bound on $E(N|k,1)$ is given by

$$(6.19)\qquad E(N|k,1) \le (k-1)/(1-\theta^2) + (1-\theta)^{-1} ,$$

<u>Proof</u>: Using the second line in (6.17) along with (6.13) gives

$$(6.20)\qquad E(N|k,1) = [(k-1)(1-\theta^2)^{-1}] + (1-\theta)^{-1} - \sum_{z=0}^{\infty} \theta^z(1-\theta^z)^{k-1}$$

which gives (6.19) directly since

$$\sum_{z=0}^{\infty} \theta^z(1-\theta^z)^{k-1} \ge 0 .$$

Corollary 6.3 $E(N|k,1)$ is a strictly increasing function of θ, $0 < \theta < 1$.

Proof: Re-write (6.20) as

$$(6.21) \qquad [(k-1)(1-\theta^2)^{-1}] - \sum_{z=0}^{\infty} \theta^z[(1-\theta^z)^{k-1}-1]$$

and differentiate to obtain

$$(6.22) \quad 2\theta(k-1)(1-\theta^2)^{-2} + (k-1)\sum_{z=0}^{\infty} z\theta^{2z-1}(1-\theta^z)^{k-2} + \sum_{z=0}^{\infty} z\theta^{z-1}[1-(1-\theta^z)^{k-1}] \, ,$$

which is positive for all $0 < \theta < 1$.

As a final corollary to Theorem 6.1, we study the rate of increase of $E(N|k,1)$ as a function of k (θ fixed again). This result is needed in the next section.

Corollary 6.4 For any $\theta (0 \le \theta \le 1)$, and $k > 2$,

$$(6.23) \qquad (k^2-1)E(N|k,1) > k^2E(N|k-1,1).$$

Proof: Using (6.20), we obtain

$$(6.24) \quad E(N|k,1) - \left(k^2/(k^2-1)\right) E(N|k-1,1)$$

$$= \frac{(k^2-k+1)}{(k^2-1)(1-\theta^2)} - \frac{1}{(k^2-1)(1-\theta)} + \sum_{z=1}^{\infty} \theta^z(1-\theta^z)^{k-2}\left[\frac{k^2}{k^2-1} - (1-\theta^z)\right]$$

$$= \frac{k(k-1)-\theta}{(k^2-1)(1-\theta^2)} + \sum_{z=1}^{\infty} \theta^z(1-\theta^z)^{k-2}\left[\frac{1}{k^2-1} + \theta^z\right] > 0 \, .$$

For the one at a time sampling plan, we assume that prior to each stage of sampling an experiment which assigns equal probability to each contender is performed to decide from which population the observation is to be taken. This assumption is merely a matter of convenience, and the results would be the same if prior to experimentation one of the k! orders of sampling were chosen at random and strict rotation were then used during the experiment - with non contenders dropped from the rotation. The non-zero "transition" probabilities are

(6.25a) $P(i,i) = P\{C(m+1) = i | C(m) = i\} = \theta$ $\qquad i = 2,\ldots,k$

(6.25b) $P(i-1,i) = P\{C(m+1) = i-1 | C(m) = i\} = \big((i-1)/i\big)(1-\theta)$ $\qquad i = 3,\ldots,k$

(6.25c) $P(1,i) = P\{C(m+1) = 1 | C(m) = i\} = (1/i)(1-\theta)$ $\qquad i = 3,\ldots,k$

(6.25d) $P(1,2) = P\{C(m+1) = 1 | C(m) = 2\} = (1-\theta)$.

Theorem 6.2 The expected number of observations for one at a time sampling when $P^* = 1$ is given by

(6.26) $\qquad E(N|k,1) = (k-1)(k+2)/2k(1-\theta)$.

Proof: Letting $S^{(k)}$ be the number of the last stage at the beginning of which there are k contenders (i.e. $C(S^{(k)}-1) = k$, $C(S^{(k)}) < k$), we have from (6.25)

(6.27a) $\qquad P(S^{(k)} \geq m) = \theta^{m-1}$

(6.27b) $\qquad P(S^{(k)} = m, C(m) = 1) = \theta^{m-1}(1-\theta)/k$

(6.27c) $\qquad P(S^{(k)} = m, C(m) = k-1) = (k-1)\theta^{m-1}(1-\theta)/k$

and from the P^* requirement, we have

(6.28) $\qquad P\left(C.I. | S^{(k)} = m, C(m) = k-1\right) = 1$.

Using the P^* requirement, (6.27) and (6.28), and recalling that a correct identification is equivalent to having $C = 1$, we obtain

$$(6.29) \quad 1 = \sum_{m=1}^{\infty} \left\{ P\left(S^{(k)}=m, C(m)=1\right) + P\left(S^{(k)}=m, C(m)=k-1\right) P\left(C.I. | S^{(k)}=m, C(m)=k-1\right)\right\}$$

$$= \sum_{m=1}^{\infty} \theta^{m-1}(1-\theta) = (1-\theta)\sum_{m=1}^{\infty} P(S^{(k)} \geq m) = (1-\theta)E(S^{(k)}|k,1) .$$

Also we have the recursive relation

$$(6.30)\quad E(N|k,1) = \sum_{m=1}^{\infty}\left\{mP\left(S^{(k)}=m, C(m)=1\right) + [m+E(N|k-1,1)]P\left(S^{(k)}=m, C(m)=k-1\right)\right\}$$

$$= \sum_{m=1}^{\infty} m\theta^{m-1}(1-\theta) + [(k-1)E(N|k-1,1)/k]\sum_{m=1}^{\infty}\theta^{m-1}(1-\theta)$$

$$= (1-\theta)^{-1} + (k-1)E(N|k-1,1)/k \ .$$

Rewriting (6.30) gives

$$(6.31)\qquad k(1-\theta)E(N|k,1) = k + (k-1)(1-\theta)E(N|k-1,1) \ .$$

Summing both sides of (6.31) over $3 \le k \le L$, and simplifying gives

$$(6.32)\qquad L(1-\theta)E(N|L,1) = \frac{L(L+1)}{2} - 3 + 2(1-\theta)E(N|2,1)$$

and using (4.2) in (6.32) with $P^* = 1$, gives (6.26).

7. General k, Vector Sampling, $P^* < 1$.

Due to the variety of possible values for $C(m)$, description of the randomization probabilities becomes more cumbersome here than for $k = 2$. We shall in fact want the probability of not taking an $(m+1)^{st}$ stage sample, which is $(1-p_m)$, to depend on $C(m)$, and we could also allow this probability to depend on the entire sequence $C(1),\ldots,C(m)$, i.e. the entire sampling history, so that the randomizing probabilities themselves become random variables. Fortunately, we shall never need a complete notation for these probabilities, and we shall use R to denote the rule by which each p_m is constructed as sampling progresses. The randomizing constant when $C(m)=k$ will be denoted simply as p_m, and this represents no loss of generality since m (when $C(m)=k$) completely characterizes the sampling history.

To avoid further complication of notation, we shall assume that if the experiment is terminated by randomization while c contenders remain, each one is given probability $(1/c)$ of being paired with F_2. Also we shall assume complete symmetry of sampling and stopping rules so that P(C.I.) and E(N) will be independent of the true pairing (Π_i, F_s), $i=1,\ldots,k$.

As in Section 6, we let $S^{(k)}$ denote the number of the last stage at which k observations are taken. Because of randomization, obtaining the distribution of $S^{(k)}$ requires some care, and we have

$$
(7.1)\qquad
\begin{aligned}
&P(S^{(k)} \geq m) = p_0 p_1 \cdots p_{m-1}\theta^{k(m-1)} \\
&P(S^{(k)} = m \mid S^{(k)} \geq m) = (1-\theta^k) + \theta^k(1-p_m) .
\end{aligned}
$$

The transition probabilities corresponding to (6.2) are more difficult to define here, and we shall consider only transitions from the k contender state. First, we define the conditional probability of having j contenders after the m^{th} sample, and before the m^{th} stage randomization, as

$$
(7.2a)\qquad
\begin{aligned}
P(k,j) &= P\Big(C(m)=j \mid C(m-1)=k,\ \text{take } m^{th} \text{ sample}\Big) \\
&= P\Big(C(m)=j \mid S^{(k)} \geq m\Big) = \binom{k-1}{k-j}\theta^j(1-\theta)^{k-j} \qquad j = 2,\ldots,k ,
\end{aligned}
$$

and the probability of stopping at the m^{th} stage when k observations are taken as

$$
(7.2b)\qquad P\Big(k(m)1\Big) = P\Big(\text{stop after } m^{th} \text{ sample} \mid S^{(k)} \geq m\Big) = (1-\theta) + \theta(1-\theta)^{k-1} + (1-p_m)\theta^k .
$$

The probability of a correct identification after the m^{th} sample is

$$
(7.2c)\qquad P(\text{C.I. at } m \mid S^{(k)} \geq m) = (1-\theta) + \theta(1-\theta)^{k-1} + (1-p_m)\theta^k(1/k).
$$

Note that in (7.2b) and (7.2c) the probabilities p_m appear with the factor θ^k, indicating that prior to stage (m+1) there were still k contenders, and probability p_m was assigned to taking the $(m+1)^{st}$ sample. In (7.2a) we avoided including p_m because at the $(m+1)^{st}$ sample there would be fewer than k contenders, p_m would depend on the number j of contenders, and we are not specifying the p_m values in this case. We avoid the need to specify these by defining the conditional probability of a correct identification given that the transition from k contenders to $2 \le j \le k-1$ contenders occurs at the m^{th} sample to include the contribution from p_m (just as the unconditional probability of correct identification includes the contribution from p_0, the probability of taking <u>any</u> sample). We write this as

$$(7.3) \qquad P\left(C.I.|k(m)j\right) = P\left(C.I.|S^{(k)} \ge m, C(m) = j\right) \qquad j = 2,\ldots,k-1 ,$$

which will be a function of the randomizing rule R which is used. Further, we shall need notation for the conditional expected additional sample size given the transition from k to j contenders at stage m (prior to the m^{th} stage randomization), and we write

$$(7.4) \qquad E\left(N_A|k(m)j\right) = E\left(N|S^{(k)} \ge m, C(m) = j, R\right) - km .$$

Unlike the situation for $k = 2$, the average sample size for general k is not completely determined by the P^* requirement. Further, we shall see that minimizing the average sample size under the P^* constraint (1.3) does not uniquely determine the sampling plan, just as meeting the P^* requirement for $k = 2$ did not uniquely determine the plan.

We denote by P(C.I.) the probability of correct identification for a given rule R, with k populations, and by $E(N|k,P^*,R)$ the average sample size for a rule R, satisfying (1.3), with k populations. We do not need to consider R when $P^* = 1$ since randomization is then impossible and we write $E(N|k,1)$ as in Section 6.

<u>Theorem 7.1</u> For vector at a time sampling, for any randomizing rule R which satisfies (1.3),

$$(7.5) \qquad E(N|k,P^*,R) \ge [(kP^*-1)/(k-1)]E(N|k,1) ,$$

with equality if and only if

$$(7.6) \qquad P(C.I.) = P^*$$

independently of the true pairing (Π_i, F_s), and the rule R does not allow randomization for stopping without an $(m+1)^{st}$ sample if $2 \le C(m) \le k-1$, or equivalently (see (7.3))

$$(7.7) \qquad P(C.I.|k(m)j) = 1 \qquad j = 2,\ldots,k-1 \text{ , all } m.$$

The requirements (7.6) and (7.7) imply the following restriction on the constants p_m (randomization probabilities when $C(m)=k$):

$$(7.8) \quad [(kP^*-1)/(k-1)](1-\theta^k)^{-1} = \sum_{m=0}^{\infty} (p_0 \cdots p_m)\theta^{km} = E(S^{(k)}|k,P^*).$$

The requirement (7.6) alone does not give equality in (7.5).

<u>Proof</u>:

First we note the following recursive relations which follow from (7.1), (7.2), (7.3) and (7.4):

$$(7.9) \quad P(C.I.) = \left(\frac{1}{k}\right)\left\{1+[k(1-\theta)+k(1-\theta)^{k-1}-k(1-\theta)^k+\theta^k-1] \sum_{m=1}^{\infty} P(S^{(k)}\ge m) \right.$$
$$\left. +k\sum_{m=1}^{\infty} P(S^{(k)}\ge m)\sum_{j=2}^{k-1} P(k,j)P(C.I.|k(m)j)\right\},$$

and

$$(7.10) \quad E(N|k,P^*,R) = k\sum_{m=1}^{\infty} P(S^{(k)}\ge m) + \sum_{m=1}^{\infty} P(S^{(k)}\ge m)\sum_{j=2}^{k-1} P(k,j)\, E\left(N_A|k(m)j\right)$$

Then we proceed by induction, assuming that the theorem is true for $j=2,\ldots,k-1$ (it has been demonstrated for $k=2$). Thus in particular

$$(7.11) \qquad E(N_A|k(m)j) \geq \left(\frac{jP(C.I.|k(m)j)-1}{j-1}\right) E(N|j,1) .$$

Substituting (7.11) into (7.10) gives a lower bound for $E(N|k,P^*,R)$, which we shall call E^*, and we have

$$(7.12) \quad E^* = k\sum_{m=1}^{\infty} P(S^{(k)}\geq m) + \sum_{m=1}^{\infty} P(S^{(k)}\geq m)\sum_{j=2}^{k-1} \frac{P(k,j)\left[jP\left(C.I.|k(m)j\right)-1\right]E(N|j,1)}{j-1}.$$

Next we find the $P\left(C.I.|k(m)j\right)$ and $\{p_m\}$ values which minimize E^* subject to requirement (7.9). We write

$$(7.13) \quad Z_k = \sum_{m=1}^{\infty} P(S^{(k)}\geq m),\ Z_j = \sum_{m=1}^{\infty} P(S^{(k)}\geq m)P\left(C.I.|k(m)j\right) \quad j = 2,\ldots,k-1.$$

We then want to find $Z_2,\ldots,Z_k$ to minimize

$$(7.14) \qquad E^* = Z_k\left\{k-\sum_{j=2}^{k-1}[P(k,j)E(N|j,1)/(j-1)]\right\}$$

$$+\sum_{j=2}^{k-1}\left(jP(k,j)E(N|j,1)Z_j\right)/(j-1)$$

subject to the requirements

$$(7.15) \quad \left(kP(C.I.)-1\right) = [k(1-\theta)+k(1-\theta)^{k-1}-k(1-\theta)^k+\theta^k-1]Z_k+k\sum_{j=2}^{k-1}P(k,j)Z_j ,$$

$$(7.16a) \qquad Z_j \geq 0 \qquad j = 2,\ldots,k ,$$

and

$$(7.16b) \qquad Z_j \leq Z_k \qquad j = 2,\ldots,k-1 .$$

Thus we proceed by induction, assuming that the theorem is true for $j=2,\dots,k-1$ (it has been demonstrated for $k=2$). Thus in particular

$$(7.11)\qquad E(N_1|x(m)j) \ge \left(\frac{\sum P(C.I.|x(m)j)-1}{j-1}\right) E(N|j,1) .$$

Substituting (7.11) into (7.10) gives a lower bound for $E(N|x^*,\beta)$, which we shall call B^*, and we have

$$(7.12)\qquad B^* = 1\cdot\sum_{m=1}^{\infty} P(S^{(1)}=m) + \sum_{m=1}^{\infty} P(S^{(k)}=m) \sum_{j=2}^{k-1} \frac{P(M=j)\left[P\left(C.I.|x(m)j\right)-1\right]E(N|j,1)}{j-1} .$$

Next we find the $P\left(C.I.|x(m)j\right)$ and $\{p_m\}$ values which minimize B^* subject to requirement (7.9). We write

$$(7.13)\qquad Z_k = \sum_{m=1}^{\infty} P(S^{(1)}=m),\quad Z_j = \sum_{m=1}^{\infty} P(S^{(k)}=m)P\left(C.I.|x(m)j\right) \qquad j=2,\dots,k .$$

We then want to find $Z_2,\dots,Z_k$ to minimize

$$(7.14)\qquad B^* = Z_k\left\{k-\sum_{j=2}^{k-1}\left[P(k,j)E(N|j,1)/(j-1)\right]\right\} + \sum_{j=2}^{k-1}\left(P(k,j)E(N|j,1)Z_j/(j-1)\right)$$

subject to the requirements

$$(7.15)\qquad \left(P^*(C.I.)-1\right) = [k(1-\theta)+k(1-\theta)]^{k-1}(1-\theta)^{k}\theta^{k}\ldots \sum_{j=2}^{k-1} \ldots$$

$$(7.16a)\qquad Z_j \ge 0, \qquad j=2,\dots,k ;$$

and

$$(7.16b)\qquad Z_k \ge Z_j, \qquad j=2,\dots,k-1 .$$

In this form we have a standard linear programming problem and it is straightforward to verify that the minimum value of E^* occurs when $Z_j = Z_k$ $(j=2,\ldots,k-1)$ if and only if

$$(7.17) \qquad \big(k/(k-1)\big)E(N|k,1) > \big(j/(j-1)\big)E(N|j,1) \qquad j = 2,\ldots,k-1.$$

The truth of condition (7.17) follows directly from (6.23). The requirement $Z_j=Z_k$ is equivalent to $P\big(C.I.|k(m)j\big) = 1$ for all m $\big($see (7.13)$\big)$ so that (7.15) becomes

$$(7.18) \qquad \big(KP(C.I.)-1\big)/(k-1)(1-\theta^k) = Z_k$$

and

$$(7.19) \qquad E^* = Z_k\left\{k + \sum_{j=2}^{k-1} P(k,j)E(N|j,1)\right\}.$$

Using (6.3b) followed by (7.18) in (7.19) gives

$$(7.20) \qquad E^* = Z_kE(N|k,1)(1-\theta^k) = \big(kP(C.I.)-1\big)/(k-1).$$

It then follows from (1.3) that E^* is lowest when (7.6) is true. It is easily verified that the bound in (7.5) is achieved by the given rule, noting that there is equality in (7.11) when $P\big(C.I.|k(m)j\big) = 1$. Under (7.6), (7.18) becomes (7.8). This concludes the proof.

Note: It is apparent that for an unsymmetric rule satisfying (1.3), the above argument can be applied for each fixed true pairing giving the bound (7.5) for E(N) under each pairing and implying that (7.6), (7.7) and (7.8) are necessary for max E(N) (max over pairings) to achieve (7.5). Thus very little "unsymmetry" can be allowed in rules achieving (7.5).

The assignment of probability (1/k) to each pairing when a randomized stop occurs is the only symmetry condition which could be relaxed, subject to a restricti similar to (3.9).

Note that no rule achieving the lower bound (7.5) can be fully truncated because of the requirement (7.7). Among the rules which do attain (7.5) we now find the one which minimizes $E(N^2)$.

Theorem 7.2 For vector at a time sampling, subject to (7.6), (7.7) and (7.8), the minimum of $E(N^2)$ is achieved by the rule which assigns

$$p_o = p_1 = \dots = p_{r_o - 2} = 1$$

$$p_{r_o - 1} = \frac{(k-1)\theta^{k(r_o-1)} - k(1-P^*)}{(1-\theta^k)\theta^{k(r_o-1)}} \tag{7.21}$$

and

$$p_{r_o} = p_{r_o+1} = \dots = 0 \ ,$$

where r_o is the smallest integer greater than or equal to

$$\frac{\log[k(1-P^*)/(k-1)]}{k \log \theta} \ . \tag{7.22}$$

Proof: Write for the expected value of N^2 given that $S^{(k)} \geq m$ and $C(m) = j$ ($j = 1,\dots,k-1$) (prior to the m^{th} stage randomization) ($m \geq 1$)

$$E\left(N^2 | k(m)j\right) , \tag{7.23}$$

and $\left(\text{see } (7.2)\right)$ write

$$E\left(N^2, (k(m)j)\right) = E\left(N^2 | k(m)j\right) P(k,j) P(S^{(k)} \geq m) \qquad 2 \leq j \leq k-1 \tag{7.24a}$$

$$E\left(N^2, (k(m)1)\right) = E\left(N^2 | k(m)1\right) P\left(k(m)1\right) P(S^{(k)} \geq m) \ . \tag{7.24b}$$

Clearly,

$$E(N^2) = \sum_{m=0}^{\infty} \sum_{j=1}^{k-1} E\left(N^2, (k(m)j)\right) . \tag{7.25}$$

Note that no rule achieving the lower bound (7.5) can be fully truncated per-
[illegible] of the requirement (7.7). Among the rules which do attain (7.5) we may find
the one which minimizes $E(N^0)$.

Theorem 7.2 For [illegible] at a time sampling, subject to (7.5), (7.7) [illegible]
the minimum of $E(N^0)$ is achieved by the rule which assigns

$$p_0 = p_1 = \cdots = p_{n_0 - 1} = 1$$

$$(7.21) \qquad p_{n_0} = \frac{(1-A)\theta^{[illegible]} - K(1-P^0)}{(1-K)\theta^{[illegible]}}$$

and

$$p_{n_0+1} = \cdots = 0,$$

where n_0 is the smallest integer greater than or equal to

$$(7.22) \qquad \frac{\log [K(1-P^0)/(1-A)]}{K \log \theta} .$$

Proof: Write for the expected value of N^0 given that $a^{(N)} = n$ and $C(N) = j$

$(j = 1, \ldots, N-1)$ [illegible] $(n \geq 1)$

$$(7.23) \qquad E\left(N^0 \mid N(n) j\right),$$

and (see (7.[illegible])) write

$$(7.24a) \qquad E\left(N^0, (N = j)\right) = [illegible] \qquad n = 2, \ldots$$

$$(7.24b) \qquad E\left(N^0, (N = 1)\right) = P\left(N^0 \mid N(n)1\right) P\left(N(n)1\right) [illegible] .$$

Clearly,

$$(7.25) \qquad E(N^0) = \sum_{n=0}^{\infty} \sum_{j=1}^{N-1} E\left(N^0, (N = j)\right) .$$

Using (7.1) and (7.2b) in (7.24b), for $m \geq 1$

$$(7.26)\quad E\left(N^2,(k(m)1)\right) = k^2m^2P(S^{(k)}\geq m)[(1-\theta) + \theta(1-\theta)^{k-1} + (1-p_m)\theta^k] .$$

Using the restriction (7.7) along with (7.1) and (7.2a) in (7.24a), we have for $m \geq 1$,

$$(7.27)\quad E\left(N^2,(k(m)j)\right) = P(S^{(k)}\geq m)P(k,j)\sum_{n=0}^{\infty}(k_m+n)^2Q(n|j), \quad j = 2,\ldots,k-1 ,$$

where $Q(n|j)$ is the density of the sample size for j populations when a $P^*=1$ requirement is imposed, so that

$$(7.28)\quad \sum_{n=0}^{\infty} Q(n|j) = 1 , \sum_{m=0}^{\infty} nQ(n|j) = E(N|j,1) , \sum_{m=0}^{\infty} n^2Q(n|j) = E(N^2|j,1) .$$

Note that by (7.8)

$$(7.29)\qquad \sum_{m=1}^{\infty} P(S^{(k)}\geq m) = (kP^*-1)/(k-1)(1-\theta^k) .$$

We put (7.26) and (7.27) in (7.25) and then use (7.28) and (7.29) to obtain

$$(7.30)\ E(N^2) = \sum_{m=1}^{\infty}(k^2m^2)P(S^{(k)}\geq m)\left\{(1-\theta)+(1-\theta)^{k-1}-(1-\theta)^k+\sum_{j=2}^{k-1}P(k,j)+\theta^k(1-p_m)\right\}$$

$$+\sum_{m=1}^{\infty}(2km)P(S^{(k)}\geq m)\left[\sum_{j=2}^{k-1}P(k,j)E(N|j,1)\right]$$

$$+\sum_{m=1}^{\infty}P(S^{(k)}\geq m)\left[\sum_{j=2}^{k-1}P(k,j)E(N^2|j,1)\right]$$

$$= k^2\sum_{m=1}^{\infty}m^2P(S^{(k)}\geq m)(1-\theta^kp_m) +$$

$$+2k\left[\sum_{j=2}^{k-1}P(k,j)E(N|j,1)\right]\sum_{m=1}^{\infty}mP(S^{(k)}\geq m)$$

$$+\left[\sum_{j=2}^{k-1}P(k,j)E(N^2|j,1)\right]\sum_{m=1}^{\infty}P(S^{(k)}\geq m)$$

[illegible]

(1.29) [illegible]

[illegible]

(1.30) [illegible]

[illegible]

(1.31) [illegible]

[illegible] $\geq T^3$

[illegible]

(1.32) [illegible]

[illegible]

$$= \left\{k^2+2k\left[\sum_{j=2}^{k-1} P(k,j)E(N|j,1)\right]+\sum_{j=2}^{k-1} P(k,j)E(N^2|j,1)\right\}\frac{(kP^*-1)}{(k-1)(1-\theta^k)}$$

$$+\left\{2k^2+2k\left[\sum_{j=2}^{k-1} P(k,j)E(N|j,1)\right]\right\}\sum_{m=1}^{\infty}(m-1)P(S^{(k)}\geq m) .$$

Note that the $\{p_m\}$ appear in (7.30) only through $\sum_{m=1}^{\infty} mP(S^{(k)}\geq m)$, which is to be minimized subject to (7.29). This problem was solved in Theorem 3.2, and (7.21) is the solution.

This completes the proof.

8. General k, one at a time sampling, $P^* < 1$.

The notation and assumptions regarding randomization probabilities, etc., are as in Section 7. As in Section 6, if an observation is to be taken, it is taken from a randomly selected contending population.

Analogous to equations (7.1) we have

$$(8.1)\qquad \begin{aligned} P(S^{(k)}\geq m) &= p_o p_1 \cdots p_{m-1}\,\theta^{m-1} \\ P(S^{(k)}=m|S^{(k)}\geq m) &= 1-\theta p_m \end{aligned}$$

and for transition probabilities given that $S^{(k)}\geq m$ we have (as in (7.2))

$$(8.2a)\qquad P(k,k-1) = P\left(C(m)=k-1|S^{(k)}\geq m\right) = (k-1)(1-\theta)/k$$

$$(8.2b)\qquad P\left(k(m)1\right) = P\left(C(m)=1 \text{ or stop by randomizing } |S^{(k)}\geq m\right)$$

$$= \left((1-\theta)/k\right) + (1-\theta)(1-p_m) .$$

Also, the probability of a correct selection is

$$(8.2c)\qquad P(\text{C.I. at } m|S^{(k)}\geq m) = \left((1-\theta)/k\right) + (1-\theta)(1-p_m)/k .$$

(Note that here, leaving state k implies entering either state (k-1) or state 1.)

<u>Theorem 8.1</u> For one at a time sampling, subject to requirement (1.3), for any randomization rule R,

$$(8.3)\quad E(N|k,P^*,R) \geq [kP^*-1)/(k-1)]E(N|k,1) = (kP^*-1)(k+2)/2k(1-\theta)\ ,$$

with equality if and only if

$$(8.4)\qquad P(C.I.) = P^*$$

independently of the true pairing (Π_i, F_2), and the rule R does not allow randomization for stopping without an $(m+1)^{st}$ sample if $2 \leq C(m) \leq k-1$, or equivalently, $\big($see (7.3) and (8.2)$\big)$ R requires that

$$(8.5)\qquad P\Big(C.I.|k(m)k-1\Big) = 1\ ,\ \text{all } m.$$

The requirements (8.4) and (8.5) imply the following restriction on the p_m (randomization probabilities when $C(m) = k$):

$$(8.6)\quad (kP^*-1)/(k-1)(1-\theta) = \sum_{m=0}^{\infty} (p_o \ldots p_m)\theta^m = E(S^{(k)}|k,P^*)\ .$$

The requirement (8.4) alone does not assure equality in (8.3).

<u>Proof</u>: From (8.1) and (8.2) it is easy to verify $\big($using (7.3)$\big)$ that

$$(8.7)\quad \Big(kP(C.I.)-1\Big) = (1-\theta)(k-1)\sum_{m=1}^{\infty} P(S^{(k)} \geq m)P\Big(C.I.|k(m)k-1\Big)\ ,$$

and that

$$(8.8)\ E(N|k,P^*,R) = \sum_{m=1}^{\infty} P(S^{(k)} \geq m)\quad \Big((k-1)(1-\theta)/k\Big)\sum_{m=1}^{\infty} P(S^{(k)} \geq m)E\Big(N_A|k(m)k-1\Big)\ ,$$

where analogously to (7.4) we define the conditional additional sample size given the transition from k to (k-1) contenders at stage m as

$$(8.9)\quad E\Big(N_A|k(m)k-1\Big) = E\Big(N|S^{(k)} \geq m, C(m)=k-1,R\Big) - m\ .$$

Again we use induction, assuming (8.3) true for (k-1) in the form

$$(8.10)\quad E\left(N_A \middle| k(m)k\text{-}1\right) \geq \left[(k-1)P\left(C.I.\middle|k(m)k\text{-}1\right)-1\right](k+1)/2(k-1)(1-\theta)\ .$$

(Theorem 4.1 established this for k = 2.)

Using (8.10) in (8.8) we have the lower bound for $E(N|k,P^*,R)$

$$(8.11)\quad E^* = [1-(k+1)/2k]\sum_{m=1}^{\infty} P(S^{(k)}\geq m) + [(k^2-1)/2k]\sum_{m=1}^{\infty} P(S^{(k)}\geq m)P\left(C.I.\middle|k(m)k\text{-}1\right).$$

Putting (8.7) into (8.11) gives

$$(8.12)\quad E^* = \left[(k+1)\left(kP(C.I.)-1\right)/2k(1-\theta)\right] + \left((k-1)/2k\right)\sum_{m=1}^{\infty} P(S^{(k)}\geq m)\ .$$

The coefficient of $\sum_{m=1}^{\infty} P(S^{(k)}\geq m)$ in (8.12) is positive so that E^* is minimized when this sum is minimized subject to (8.7). Since each $P\left(C.I.\middle|k(m)k\text{-}1\right) \leq 1$, from (8.7) we have that

$$(8.13)\quad \left[\left(kP(C.I.)-1\right)/(1-\theta)(k-1)\right] = \sum_{m=1}^{\infty} P(S^{(k)}\geq m)P\left(C.I.\middle|k(m)k\text{-}1\right) \leq \sum_{m=1}^{\infty} P(S^{(k)}\geq m)\ ,$$

with equality if and only if each $P\left(C.I.\middle|k(m)k\text{-}1\right) = 1$. Using (8.13) in (8.12), and then using (1.3), gives

$$(8.14)\quad E^* \geq \left[\left(kP(C.I.)-1\right)(k+2)\right]/2k(1-\theta)$$

$$\geq (kP^*-1)(k+2)/2k(1-\theta)\ .$$

This gives (8.3), and the rest of the theorem follows from the equality in (8.10) and (8.13) when all $P\left(C.I.\middle|k(m)k\text{-}1\right) = 1$, and the equality in (8.14) when $P(C.I.)=P^*$. Under these conditions (8.7) becomes (8.6). This completes the proof.

For a discussion of the effect of non-symmetry, see the remark following Theorem 7.1.

The next theorem gives the procedure which minimizes $E(N^2)$ while achieving equality in (8.3).

Theorem 8.2 For one at a time sampling, subject to (8.4), (8.5) and (8.6), the minimum of $E(N^2)$ is achieved by the rule which assigns

$$p_0 = p_1 \quad \cdots = p_{r_1-2} = 1$$

$$(8.15) \qquad p_{r_1-1} = [(kP^*-1)-(k-1)(1-\theta^{r_1-1})]/(k-1)(1-\theta)\theta^{r_1-1}$$

$$p_{r_1} = p_{r_1+1} = \cdots = 0 ,$$

where r_1 is the smallest integer greater than or equal to

$$(8.16) \qquad \{\log[k(1-P^*)/(k-1)]\}/\log \theta .$$

The proof is similar to that of Theorem 7.2 and is omitted.

9. Comparison of Rules, General k

In order to compare the expected sample sizes for the best vector at a time and one at a time sampling plans, first we shall obtain bounds for the difference in average sample sizes when $P^*=1$.

Lemma 9.1 For any $k \geq 2$, and any $0 \leq \theta \leq 1$, let $E_v(N|k,1)$ and $E_o(N|k,1)$ be given by (6.4) and (6.26) respectively. Then

$$(9.1) \qquad [(k^2-k+2)/4k] \leq E_v(N|k,1) - E_o(N|k,1) \leq [(2-\theta)(k^2-k+2)/4k]$$

with equality at the lower bound for $\theta = 1$ and at the upper bound for $\theta = 0$.

Proof: For k=2, (9.1) is easily verified. From the inductive relation (6.3b), with (6.26) and (6.2a), we obtain for general k

$$E_v(N|k,1)-E_o(N|k,1)=(1-\theta^k)^{-1}\Big[k+\sum_{j=2}^{k-1}P(k,j)\Big\{[E_v(N|j,1)-E_o(N|j,1)]+E_o(N|j,1)\Big\}\Big]-E_o(N|k,1)$$

(9.2)

$$= (1-\theta^k)^{-1}\Big[k+\sum_{j=2}^{k-1}\binom{k-1}{j-1}\theta^j(1-\theta)^{k-j}\Big\{[E_v(N|j,1)-E_o(N|j,1)]+\frac{(j-1)(j+2)}{2j(1-\theta)}\Big\}\Big]-\frac{(k-1)(k+2)}{2k(1-\theta)}$$

$$= (1-\theta^k)^{-1}\sum_{j=2}^{k-1}\binom{k-1}{j-1}\theta^j(1-\theta)^{k-j}[E_v(N|j,1)-E_o(N|j,1)]$$

$$+ [2k(1-\theta)(1-\theta^k)]^{-1}\Big\{2k^2(1-\theta)-(k-1)(k+2)(1-\theta^k)+k(k-1)\theta^2(1-\theta^{k-2}) +$$

$$+ 2k\theta[(1-\theta^{k-1})-(1-\theta)^{k-1}]-2[(1-\theta^k)-(1-\theta)^k-k\theta(1-\theta)^{k-1}]\Big\}$$

$$= (1-\theta^k)^{-1}\Big\{\frac{(k-1)(1-\theta)}{2}+\frac{(1-\theta)^{k-1}}{k}+\sum_{j=2}^{k-1}\binom{k-1}{j-1}\theta^j(1-\theta)^{k-j}[E_v(N|j,1)-E_o(N|j,1)]\Big\}.$$

Now assume (9.1) is true for $j=2,\ldots,k-1$. Substituting the lower bound

$$\Big[(j^2-j+2)/4j\Big]$$

into the sum which appears in (9.2) gives

(9.3) $$E_v(N|k,1)-E_o(N|k,1)\geq(1-\theta^k)^{-1}\Big\{\frac{(k-1)(1-\theta)}{2}+\frac{(1-\theta)^{k-1}}{k}$$

$$+\frac{k(k-1)(\theta^2-\theta^k)+2[(1-\theta^k)-(1-\theta)^k-k\theta(1-\theta)^{k-1}]}{4k}\Big\}$$

$$= \Big(4k(1-\theta^k)^{-1}\Big)\Big\{(k^2-k+2)(1-\theta^k)+(k-1)(1-\theta)^2[k-2\theta(1-\theta)^{k-3}]+2(1-\theta)^{k-1}\Big\}$$

$$\geq (k^2-k+2)/4k\,.$$

This proves the lower bound. Equality at $\theta = 1$ follows from equality in the first line of (9.3) by induction, and in the third line of (9.3) since

[illegible]

[illegible]

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

[illegible]

[illegible]

$$\lim_{\theta\to 1} [(1-\theta)^j/(1-\theta^k)] = 0 \ , \ j \geq 2,$$

The upper bound is obtained similarly. This completes the proof.

An upper bound for $\left(E_v(N|k,1)-E_0(N|k,1)\right)$ is obtainable by using (6.19). This gives

$$(9.4) \quad E_v(N|k,1)-E_0(N|k,1) \leq [(k^2-k+2)(1-\theta)+4\theta]/2k(1-\theta^2) \ .$$

The bounds in (9.1) and (9.4) agree for $\theta = 0$; (9.1) is sharper for all θ if $k \leq 4$, and for $k > 4$, (9.4) is sharper for $\theta < \theta_k$ and (9.1) is sharper for $\theta \geq \theta_k$, where

$$(9.5) \qquad \theta_k = 1-(2/\sqrt{k}).$$

The bound (9.4) is extremely poor in the vicinity of $\theta = 1$.

Using (9.1) and (9.4), dividing by $E_0(N|k,1)$ throughout, using (6.26) and simplifying gives the following

$$(9.6) \quad 1 + [(k^2-k+2)(1-\theta)/2(k-1)(k+2)] \leq \left(E_v(N|k,1)/E_0(N|k,1)\right) \leq M(k,\theta)$$

where

$$(9.7) \quad M(k,\theta) = \begin{cases} [2k(k+\theta)/(k-1)(k+2)(1+\theta)] & , \ \theta < 1-(2/\sqrt{k}) \\ [4k^2-\theta(3-\theta)(k^2-k+2)]/2(k-1)(k+2) & , \ \theta \geq 1-(2/\sqrt{k}) \ , \end{cases}$$

again with equality below at $\theta = 1$ and above at $\theta = 0$.

Using (7.5) and (8.3) to relate average sample sizes for general P^* to those for $P^* = 1$, we obtain the following:

<u>Lemma 9.2</u> For any $k \geq 2$, any $(1/k) < P^* \leq 1$, and any $0 \leq \theta \leq 1$ with $E_v(N|k,P^*)$ given by (7.5) and $E_0(N|k,P^*)$ given by (8.3),

$$(9.8) \quad \frac{(kP^*-1)(k^2-k+2)}{4k(k-1)} \leq E_v(N|k,P^*)-E_0(N|k,P^*) \leq [(kP^*-1)L(k,\theta)]/(k-1)$$

[illegible]

[illegible]
[illegible]
[illegible]

[illegible]
[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

[illegible]
[illegible]

$$[illegible]$$

[illegible]

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]
[illegible]

$$[illegible]$$

with equality below at $\theta = 1$ and above at $\theta = 0$, where

$$(9.9) \qquad L(k,\theta) = \begin{cases} [(k^2-k+2)(1-\theta)+4\theta]/2k(1-\theta^2) & , \theta < 1-(2/\sqrt{k}) \\ (2-\theta)(k^2-k+2)/4k & , \theta \geq 1-(2/\sqrt{k}) . \end{cases}$$

Also,

$$(9.10) \qquad \frac{E_V(N|k,P^*)}{E_0(N|k,P^*)} = \frac{E_V(N|k,1)}{E_0(N|k,1)} .$$

From Lemma 9.2 we can conclude that even for $\theta = 1$, when both expected sample sizes are infinite, they differ in magnitude by at most $(k/4)$ for any P^*, while the difference never exceeds $(k/2)$ for any θ or P^* values. When θ is small, the one at a time plan is somewhat more efficient, but never reduces the expected sample size by more than $(1/2)$ for any θ, k or P^* as seen from (9.10) and (9.1).

Again if the cost of sampling is more closely tied to the number of stages than to sample size, the vector plan would be preferred. For example, using (6.18) and (6.26), the expected number of stages when $P^* = 1$ can be compared. Using (6.17) to obtain an upper bound of $(1-\theta)^{-1}$ for $E_V(M|k,1)$ we have

$$(9.11) \qquad E_0(M|k,1) - E_V(M|k,1) \geq (k^2-k-2)/2k(1-\theta)$$

with equality at $\theta = 0$, which gives the smallest value taken on in (9.11). For small k values, (9.11) tends to understate the difference. For example, when $k=2$, (9.11) gives a value of 0 whereas the true difference is $\left(\theta/(1-\theta^2)\right)$.

We shall compare the sequential procedures with the single sample procedure which takes R vectors of k observations, $(X_{1j},\ldots,X_{kj})$ $(j=1,\ldots,R)$ and if some $X_{ij} \epsilon S_2-S_1$ associates Π_i with F_2 or if $X_{ij} \epsilon S_1-S_2$ for $(k-1)$ different i indices, associates the remaining Π_i with F_2. If for $(k-c)$ indices of i, there is an $X_{ij} \epsilon S_1-S_2$, $(c\geq 2)$, and no $X_{ij} \epsilon S_2-S_1$, then the pairing is chosen at random from the remaining c possible pairings.

For this procedure to satisfy the P* requirement (1.3), R is given by

$$(9.12)\qquad k(1-P^*) = k\theta^R + (1-\theta^R)^k - 1 \,.$$

It is easily seen that

$$(9.13)\qquad R > \log (1-P^*)/\log \theta,$$

so that the sample size n is

$$(9.14)\qquad n = kR > k \log (1-P^*)/\log \theta.$$

Note that no single sample procedure can give $P^* = 1$. Using (9.14) and (8.3) we can compare the single sample and one at a time plans for average sample size, obtaining

$$(9.15)\quad [n/E_0(N|k,P^*)] > [2k^2(1-\theta)\log(1-P^*)]/(kP^*-1)(k+2)\log \theta$$

$$> 2k^2/\left[(k+2)(k-1)\left(1+\left((1-\theta)/2\right)+\left((1-\theta)^2/3\right)+\dots\right)\right].$$

This last expansion is not useful for θ near 0, but it is easily seen that the middle term of (9.15) is approximately $[2k^2/(k+2)(k-1)]$ for θ near 0. Thus for large k, the single sample procedure takes on the average approximately twice the number of observations taken by the one at a time sequential procedure.

Acknowledgement

The authors wish to thank Mr. Norman Starr for several helpful discussions in obtaining Theorem 6.1. Special thanks are also due to Professor I. R. Savage for reading an early draft and making several constructive suggestions.

References

[1] Bechhofer, R., Kiefer, J., Sobel, M. (1966). Sequential Ranking and Identification Procedures. Chicago University Press, I.M.S. Monograph Series.

[2] Kemeney, J. G., and Snell, J. L. (1959) Finite Markov Chains. Van Nostrand, Princeton.